

EYE MOVEMENTS IN HUMAN PERFORMANCE MODELING OF SPACE SHUTTLE OPERATIONS

Michael Matessa and Roger Remington
NASA Ames Research Center
Moffett Field, CA

The goal of our research is to easily develop models that predict astronaut performance in space shuttle operations, but it is difficult to make extrapolations from astronaut training data. A solution is to decompose a complex task into a set of basic operators which are sequenced to create longer chains of behavior. In this modeling project, gaze durations and sequences are predicted and compared to the performance of novice (trained pilots) and expert (astronaut) space shuttle operators. The model makes generally good zero-parameter predictions of gaze durations, but there are notable discrepancies. The gaze sequence of the model is more similar to expert performance than novice performance, but there are differences from both. It appears that with more training, experts develop different gaze sequence strategies than novices due to familiarity with fault messages and procedures. Future modeling efforts should have their gaze sequence strategies based on expert performance.

INTRODUCTION

The goal of our research is to easily develop models that predict astronaut performance in space shuttle operations. Empirical extrapolations from astronaut training data are difficult, especially for novel problems that have not been extensively trained. A solution is to decompose a complex task into a set of basic operators such as reading a phrase, pressing a key, or throwing a switch. These operators can be assigned performance parameters (validated by training data) and be sequenced to create longer chains of behavior. This method of cognitive task analysis is used by the GOMS (Card, Moran, and Newell, 1983) modeling methodology and has proven useful in representing the procedural knowledge that characterizes tasks in many domains. The model built for this paper was created with an automated version of GOMS, Apex-CPM, that allows the expression of hierarchical goal structure as a nested set of procedures, with the lowest procedures being basic operators (John et al., 2002; Vera et al., in press).

The task of interest for this paper is fault management in the space shuttle during ascent. Fault management is trained as a well-specified pattern: being alerted to a fault, identifying it, determining the correct procedure, taking actions to correct it, and verifying that the fault has been correctly managed. Since most of the time involved in fault management is a result of visually acquiring information, the modeling has focused on the prediction of gaze durations and sequences measured by eye tracking. Eye tracking has been useful in understanding human performance issues in aviation (Fitts, Jones, and Milton, 1950; Bellenkes, Wickens, and Kramer, 1997; Anders, 2001), but to date no eye tracking studies have been performed in the space shuttle environment.

The Space Shuttle Cockpit Simulator at the Intelligent Spacecraft Interface Systems (ISIS) lab at NASA Ames Research Center permits the collection of eye tracking information during shuttle operations. The simulator has been

used to obtain data from novice shuttle operators (specially trained airline pilots) and expert operators (current astronauts). The hypotheses of this modeling study are that both airline pilots and astronauts are skilled operators of complicated flying equipment, that both have similar performance parameters for basic operators, and that strategic behavior sequences can transfer between pilots and astronauts (since both are skilled at the behavioral sequences involved in following emergency procedures from checklists). Parameters for the duration of eye fixations were obtained from existing literature with similar textual material.

GOMS gaze duration prediction

The GOMS methodology was used by Chuah, John, and Pane (1994) to predict times for performing tasks using graphic and textual displays. Their model of comprehending visual information from a single fixation is constructed from an attend-target operator lasting 50 msec, an initialize-eye-movement operator lasting 50 msec, an eye-movement operator lasting 30 msec, a perceive-target operator lasting 290 msec, and a verify-target operator lasting 50 msec. This gives a total time of 470 msec per fixation. With their assumption that a fixation can encompass roughly 6 letters in 12-point font, the times for gaze durations during a particular fault in the shuttle environment can be predicted as follows: reading a key on the keyboard requires one fixation giving a gaze duration of 470 msec, reading a fault message requires two fixations giving a gaze duration of 940 msec, reading data or a switch label requires three fixations taking 1410 msec, and reading a procedure requires eleven fixations taking 5170 msec. See Table 1 for details. These predictions were compared to eye movement data collected for novice and expert shuttle operators. The sequence of gazes and manual actions needed by the model to solve shuttle malfunctions were determined from two pilots in the novice group and compared to the rest of the novices and the experts.

Keyboard = (ACK)
 = 470 * 1 = 470 ms

Message = (MPS LH2) + (/OH2 ULL)
 = 470 * 2 = 940 ms

Data = (25.7↓) + (25.6↓) + (25.8↓)
 = 470 * 3 = 1410 ms

Switch = (LH2 ULLAGE) + (PRESS) + (OPEN)
 = 470 * 3 = 1410 ms

Procedure = If 2(3)
 Ps<28.0
 or>34.0:
 MPS LH2
 ULL PRESS
 -- OP
 When all
 Ps>34.0:
 MPS LH2
 ULL PRESS
 -- AUTO
 = 470 * 11 = 5170ms

Table 1: Predictions for reading times

METHOD

Gaze duration is defined as the total time spent looking in region of interest. This may be made up of a number of individual eye fixations. The number of gazes to a region of interest is defined as the number of initial fixations to a region of interest starting from another region of interest (this does not count individual fixations within a region of interest).

The independent variables are region of interest in the shuttle cockpit (message, data, keyboard, procedures, or switch) and agent (pilots, astronauts, or model). The dependent variables are average gaze duration and average number of gazes.

Participants

Five airline transport pilots with an average of 15,000 flight-hours experience participated in the novice condition. Five current astronauts with a minimum of two years of training participated in the expert condition.

Apparatus

The Space Shuttle Cockpit Simulator at the Intelligent Spacecraft Interface Systems (ISIS) lab at NASA Ames Research Center was used for the experiment. The simulator is a fixed-base, part-task simulator with 4 20" LCD monitors for the 7 front displays, 7 20" touch-screen LCD monitors for the side and overhead switch panels, 1 12" touch-screen LCD monitor for the keyboard, and 6 audio speakers. The shuttle flight dynamics and system parameter tables were provided by the Shuttle Engineering Simulator at Johnson Space Center. The display graphics were generated with the Virtual Prototypes Incorporated's Visual APplicationS builder (VAPS), a C-based rapid prototyping tool. A head-mounted eye camera (ISCAN ETL-500, ISAN, Inc., Burlington, MA) and head tracker (FasTRAK, Polhemus, Colchester, VT) were used to measure the participants' eye movements.

Procedure

Prior to the simulation runs, participants in the novice condition participated in a 1-week training course. Participants in both novice and expert conditions were given a simulator familiarization session. Each trial simulated the ascent phase of shuttle operations, starting at launch and ending at Main Engine Cut-Off (MECO) which occurs at 8 minutes and 30 seconds in simulated mission elapsed time. During the trial a main engine malfunction was inserted. The Flight Data File procedure checklist, which lists all the steps required to recover from the malfunctions, was provided to the participants during the simulation. Analyses on gazes in regions of interest relevant to solving the malfunction were performed beginning with the fault alarm and ending with the first manual response for the malfunction.

Results

Regions of interest in the cockpit were chosen that were relevant to reacting to the malfunction, and the average duration of participants' gazes to the regions of interest were measured. The region corresponding to the procedure checklist did not have a visual plane specified by the eye tracker, and so gazes in that region were regarded as having "no specific plane". Gazes looking away from the displays in any other direction were also regarded as having no specific plane, but an informal review of the data showed that gazes in no specific plane that lasted more than three seconds were directed to the procedure checklist. This assumption should be verified in future investigations. The model predictions of gaze durations and gaze sequence were tested against the durations for three pilots randomly chosen to test the model and the four astronauts who looked at the regions of interest (one astronaut anticipated the malfunction alarm and solved the malfunction with few fixations).

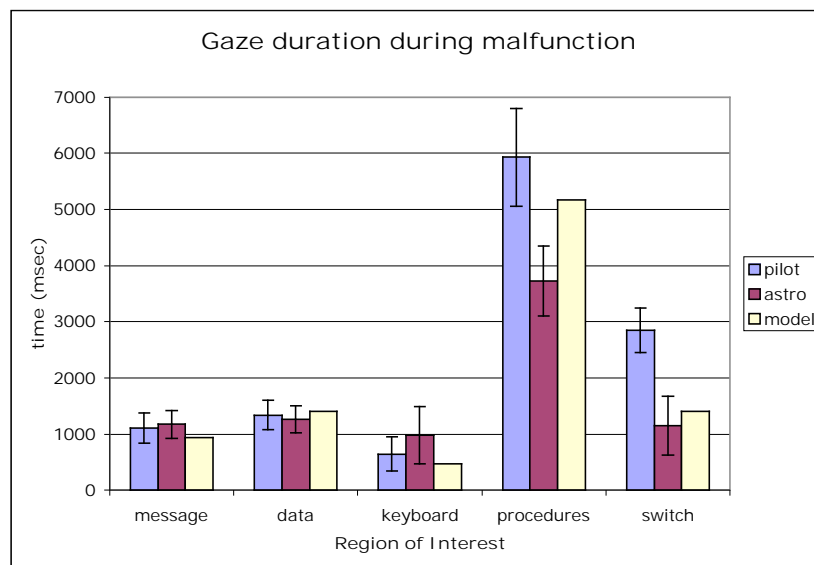


Figure 1: Average gaze durations for regions of interest.

Model creation

The model of shuttle malfunction response consists of a sequence of gazes and motor actions. The durations of the gazes are predictions based on the GOMS analysis above. The sequence of the gazes was determined from an analysis of a subset of the pilots. This allows the model to make predictions for the pilots not used in the model analysis and the astronauts. Two pilots were randomly chosen and their gaze and motor sequences were analyzed to create a sequence for the model. Both pilots showed sequences that included extra gazes to verify information they had already looked at. These extra gazes were removed to create a minimal sequence of gazes and manual responses necessary to solve a single malfunction (Table 2). This minimal sequence was expected to be more representative of astronaut behavior.

```
(procedure
(index
(respond to malfunction))
(read message)
(read keyboard)
(press keyboard)
(read procedure)
(read message)
(read data)
(read procedure)
(read switch)
(throw switch))
```

Table 2: Model description of sequence.

The sequence consisted of reading the fault message, reading the label of a key that acknowledges the fault message, pressing the key, reading the first part of the malfunction procedure that indicates what data to check, verifying the fault message, reading the data, reading the

second part of the procedure that indicates which switch to throw, reading the switch label, and throwing the switch. Again, the model predicts times for reading based on the GOMS analysis above. The reading times for particular regions of interest were expected to be similar for pilots and astronauts, but

Model evaluation

Figure 1 shows the gaze duration predictions of the model tested against data from the pilots and astronauts, with error bars indicating the standard error. The average difference of gaze duration between the model and pilots for the five regions of interest was 523ms. The correlation between the gaze durations for the regions of interest of the model and pilots was .96, and the correlation between the gaze durations of the model and astronauts was .99. However, these large correlations are mostly driven by the large difference between the procedure gaze duration and other gaze durations. The average difference between the model and astronauts was 520ms. The average percent error between pilot and model was 22%, and the average percent error between astronaut and model was 29%.

Example gaze sequences for a pilot, an astronaut and the model can be seen in Figure 2. Horizontal lines represent the time spent looking at particular regions of interest. Gaps in the lines represent time spent on regions that were not directly relevant to solving the malfunction. The figure shows how the pilot's verification of information results in an increased time to solve the malfunction compared to the model based on minimal gazes. The time for the model to solve the malfunction is closer to the time for the astronaut, but differences include a longer time for the astronaut to look at the data and the astronaut looking at the procedure only at the end of the sequence.

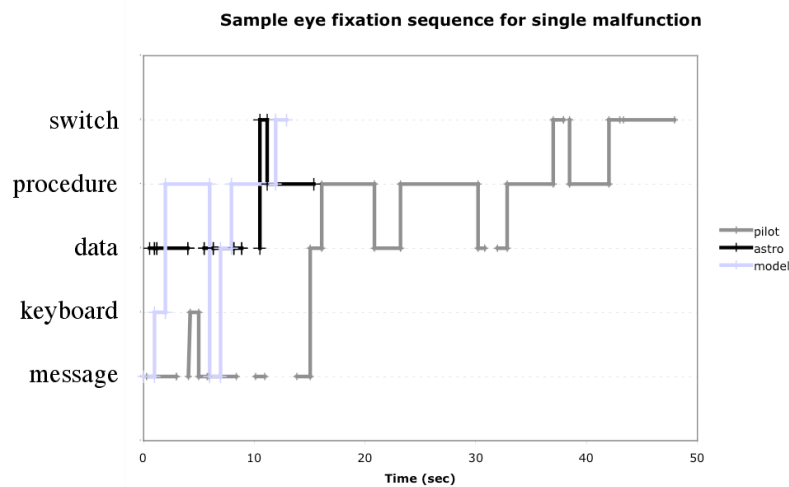


Figure 2: Example eye gaze sequences.

With the large individual differences in gaze sequences, it is difficult to easily compare the sequences of the model, pilots, and astronauts. One simplification is to analyze the number of gazes at a particular region of interest. From the examples in Figure 2 it can be seen that this metric can show trends such as pilots having more gazes at the procedure and data regions than the model, or astronauts having fewer gazes at the procedure region than the model. The number of gazes for each relevant region of interest for pilots, astronauts, and the model can be seen in Figure 3. Again, error bars in the figure show the standard error for the three pilots and four astronauts.

Pilots are seen to look at the message and procedure regions more often than the astronauts. Pilots look at the message and switch regions more often than the model. The model looks at the message and procedure regions more often than astronauts, and the model looks at the data region less often than astronauts. The average difference of number of gazes between the model and pilots for the five regions of interest was 1.1 gazes. The average difference between the model and astronauts was 0.8 of a gaze. The correlation between the number of gazes for the regions of interest of the model and pilots was .67. The correlation between the number of gazes of the model and astronauts was -.67. The average percent error between pilot and model was 41%, and the average percent error between astronaut and model was 47%.

Discussion

The model makes generally good zero-parameter predictions of gaze durations for pilots and astronauts, but there are notable discrepancies. Gaze duration in the model represents the process of identifying visual information, but does not include the time to search for the information. The increased gaze time on the switch for the pilots relative to the model may represent some visual search time, as the switch panel contains many similar-looking switches. The model also assumes that all information at a location is processed once

and only once. The decreased gaze time on procedures for the astronauts relative to the model may represent the astronaut reading only a particular part of the procedure. Likewise, the increased gaze time on procedures for the pilots relative to the model may be due to the pilots reading parts of the procedure more than once.

The gaze sequence of the model was based on simplified strategies used by the training pilots, and was similar to the performance of the astronauts, differing in the average number of gazes to a particular region of interest by 0.8 of a gaze. However, the astronauts have more gazes at the data and fewer at the fault message and procedures. It appears that with more training, astronauts develop different gaze sequence strategies than pilots due to familiarity with fault messages and procedures. Future modeling efforts should have their gaze sequence strategies based on astronaut performance.

The analyses in this paper include only regions of interest that are relevant to solving malfunctions, but there is also a parallel task of monitoring the entire state of the shuttle. Future research will be directed towards modeling the multitasking behavior of the astronauts as they solve malfunctions while at the same time monitor the overall state of the shuttle.

ACKNOWLEDGEMENTS

This work was supported by an Office of Naval Research grant, a NASA Space Human Factor Engineering grant, and the Airspace Operations System project of NASA's Airspace System program.. The members of the ISIS lab were instrumental in participant training and data collection. ISIS members include Steven Elkins, Valerie Huemer, Bob Lawrence, Jeff McCandless, Rob McCann, and Fritz Renema.

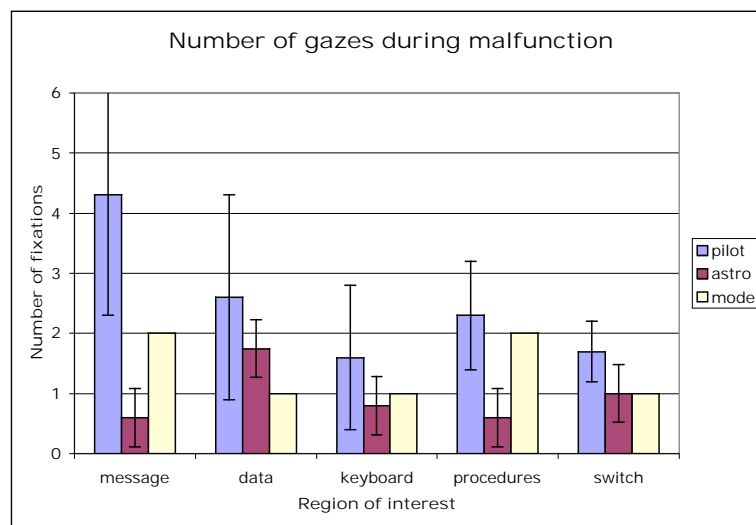


Figure 3: Number of gazes for regions of interest.

REFERENCES

- Anders, G. (2001). Pilots attention allocation during approach and landing: eye and head tracking research in an A330 full flight simulator. In R. Jensen (Ed.), *Proceedings of the 11th International Symposium on Aviation Psychology*. Columbus: Ohio State University.
- Bellenkes, A., Wickens, C.D., & Kramer, A. (1997). Visual scanning and pilot expertise: The role of attentional flexibility and mental model development. *Aviation, Space, and Environmental Medicine*, 68 (7).
- Card, S. K., Moran, T.P. & Newell, A. (1983). *The Psychology of Human-Computer Interaction*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Chuah, M.C., John, B.E., and Pane, J. "Analyzing Graphic and Textual Layouts with GOMS: Results of a Preliminary Analysis," *CHI 94 Conference Companion: Conference on Human Factors in Computing Systems*, Boston, MA, April 1994, pp. 323-324.
- Fitts, P. M., Jones, R. E. & Milton, J. L. (1950). Eye movements of aircraft pilots during instrument-landing approaches. *Aeronautical Engineering Review*, 9(2), 24-29.
- John, B. E., Vera, A. H., Matessa, M., Freed, M., and Remington, R. (2002) Automating CPM-GOMS. In *Proceedings of CHI'02: Conference on Human Factors in Computing Systems*. ACM, New York.
- Vera, A., John, B., Remington, R., Matessa, M., & Freed, M. (in press). Automating Human-Performance Modeling at the Millisecond Level. *Human-Computer Interaction*.

This work is not subject to U.S. copyright restrictions.